

Appendix A

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 13, 2005, 09:51:41 ; Search time 3878.11 Seconds
(without alignments)
9982.023 Million cell updates/sec

Title: US-10-511-270-3
Perfect score: 1017
Sequence: 1 cgggatccatgctgggcccc.....tgagctgtctcagaattccg 1017

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 34239544 seqs, 19032134700 residues

Total number of hits satisfying chosen parameters: 68479088

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*
1: gb_est1:*
2: gb_est2:*
3: gb_htc:*
4: gb_est3:*
5: gb_est4:*
6: gb_est5:*
7: gb_est6:*
8: gb_gss1:*
9: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	921.2	90.6	1635	3	AK002457	AK002457 Mus muscu
2	908.6	89.3	1596	3	AK010857	AK010857 Mus muscu

ALIGNMENTS

RESULT 1

AK002457 1635 bp mRNA linear HTC 03-APR-2004
LOCUS Mus musculus adult male kidney cDNA, RIKEN full-length enriched library, clone:0610010D20 product:hypothetical Aminoacyl-transfer RNA synthetases class-II/Dihydrodipicolinate synthetase containing protein, full insert sequence.
DEFINITION
ACCESSION AK002457
VERSION AK002457.1 GI:12832454
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1
AUTHORS Carninci, P. and Hayashizaki, Y.
TITLE High-efficiency full-length cDNA cloning
JOURNAL Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
REFERENCE 2
AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE 20499374
PUBMED 11042159

FEATURES
source 1. 1635
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="FANTOM DB:0610010D20"
/db_xref="taxon:10090"
/clone="0610010D20"
/sex="male"
/tissue_type="kidney"
/clone_lib="RIKEN full-length enriched mouse cDNA library"
/dev_stage="adult"
CDS 68. 1033
/note="unnamed protein product; hypothetical
Aminoacyl-transfer RNA synthetases
class-II/Dihydrodipicolinate synthetase containing protein
(InterPro|IPR002106, InterPro|IPR002220, evidence:
InterPro)
putative"
/codon_start=1
/protein_id="BAB22114.1"
/db_xref="GI:12832455"
/translation="MLGPQIWASMRQGLSRGLSRNVKGGKVDIAGIYPPVTTPTATA
EVDYDKLEENLNRLATFPFRGFVVGSTGEFPFLTSLELVVSRVRQAIPKDKFLIA
GSGCESTQATVEMTVSMAQVGADVAMVTPCYRGRMSSAALIHHTKVADVSPV
LYSVPANTGLELPVDAVVTLSQHPNIIGLKDSGGDVTRIGLIVHKTSKQDFQVLGSA
GFLASYAVGAVGGICGLANVLGAQVCQLERLCLTGQWEAAQELQHRLLIEPNTAVTRR
FGIPGLKKTMDWFGYGGPCRAPLQELSPTEEEALRLDPSNNGWL"

ORIGIN
Query Match 90.6%; Score 921.2; DB 3; Length 1635;
Local Similarity 94.7%; Pred. No. 1.8e-234;

Matches 953; Conservative 0; Mismatches 53; Indels 0; Gaps 0;

Qy	9	ATGCTGGGCCCCAAATCTGGGCTCCATGAGGCAGGGCTGAGCAGGGGCTTGTCTAGG	68
Db	68	ATGCTGGGCCCCAAATTTGGGCTCCATGAGGCAGGGCTGAGCAGGGGCTTGTCTAGG	127
Qy	69	AACGTGAAGGGGAAGAAGATAGACATTGCCGGCATCTACCCACCCGTGACCACCCCATTC	128
Db	128	AATGTGAAGGGCAAGAAGGTAGACATTGCCGGCATCTACCCACCCGTGACCACCCCATTC	187
Qy	129	ACCGCCACCGCAGAAGTAGACTATGGGAAACTGGAAGAGAACCTGAACAAACTGGCCGCC	188
Db	188	ACCGCCACCGCAGAGGTAGACTATGGGAAACTGGAAGAGAACCTGAACAGACTGGCCACC	247
Qy	189	TTCCCTTTTCGAGGCTTCGTGGTCCAGGGCTCTACTGGAGAGTTTCCATTCTGACCAGC	248
Db	248	TTCCCTTTTCGAGGCTTCGTGGTCCAGGGTTCGACTGGAGAGTTTCCGTTCTGACCAGC	307
Qy	249	CTTGAGCGCCTAGAGGTGGTGAAGCCAGTGCAGGCCATACCCAAGGACAAGCTCCTG	308
Db	308	CTCGAGCGCCTGAGGTGGTGAAGCCAGTGCAGGCCATACCCAAGGACAAGTTCTG	367
Qy	309	ATAGCCGGCTCTGGCTGCGAGTCCACGCAAGCCACAGTAGAGATGACTGTCAGCATGGCT	368
Db	368	ATAGCCGGCTCTGGCTGCGAGTCCACGCAAGCCACAGTAGAGATGACTGTCAGCATGGCT	427
Qy	369	CAGGTGGGTGCTGATGCGCCATGGTGGTGACCCCTTGTTACTATCGCGGCCGATGAAC	428
Db	428	CAGGTGGGTGCTGATGTCGCATGGTGGTGACCCCTTGTTACTATCGTGGCCGATGAGC	487
Qy	429	AGCGCTGCCCTCATTACCACTACACCAAGGTTGCTGATCTTTCTCCAATCCCGTGGTG	488
Db	488	AGCGCTGCCCTCATTACCACTACACCAAGGTTGCTGACGTTTCTCCAATCCCTGTGGTG	547
Qy	489	CTGTACAGTGTCCCAGGCAACACGGGTCTAGAGCTGCCTGTGGATGCCGTGGTCACATTG	548
Db	548	TTGTACAGTGTCCCAGGCAATACGGGGCTAGAGCTACCTGTGGATGCCGTGGTTACATTG	607
Qy	549	TCTCAGCACCCAAATATCATTTGGCTTGAAGGACAGTGGTGGAGATGTGACCAGGACTGGG	608
Db	608	TCTCAGCACCCAAATATCATCGGCTTGAAGGACAGTGGTGGAGATGTGACCAGGATTGGA	667
Qy	609	CTGATTGTTTACAAGACCAGCAAGCAGGATTTCCAGGTGTTGGCTGGGTGAGTTGGCTTC	668
Db	668	CTGATAGTTTACAAGACCAGCAAGCAGGATTTCCAGGTGTTGGCTGGGTGAGTTGGCTTC	727
Qy	669	CTCCTGGCCAGCTATGCTGTGGGAGCTGTTGGGGGCATATGTGGCTGGCCAATGTCTTG	728
Db	728	CTCCTGGCCAGCTATGCTGTGGGAGCTGTTGGGGGCATATGTGGCTGGCCAATGTCTTG	787
Qy	729	GGGGCCAGGTGTGCCAGCTGGAGAGACTCTGCCTCACAGGGCAGGGGGAAGCTGCCCAG	788
Db	788	GGGGCCAGGTGTGCCAGCTGGAGAGACTCTGCCTCACAGGGCAGTGGGAAGCTGCCCAG	847
Qy	789	AGACTGCAGCACCGCTCATCGAGCCCAACACTGCGGTGACCCGGCGCTTTGGAATACCA	848
Db	848	GAACTACAGCACCGTCTCATCGAGCCCAACACTGCGGTGACCCGGCGCTTTGGAATACCA	907
Qy	849	GGGCTGAAGAAAACCATGGACTGGTTTGGCTACTATGGAGGTCCCTGCCGTGCCCCCTTG	908
Db	908	GGGCTGAAGAAAACCATGGACTGGTTTGGCTACTATGGAGGTCCCTGCCGTGCCCCCTTG	967
Qy	909	CAGGAGTTGAGCCCCCTCAGAGGAAGAGGCGCTTCGCTTGGATTTTCAACAACATGGCTGG	968
Db	968	CAGGAGCTGAGCCCCACAGAGGAGGAGGCACTGCGCTTGGATTTTCAACAACATGGCTGG	1027
Qy	969	CTTTAATGACAAGCGGGGACACCTGGTCTGAGCTGTCTCAGAATT	1014
Db	1028	CTTTAATGACAAGCAGGAGACGCTGGCTGAGCTATCTGGGACTT	1073

Result 2

LOCUS AK010857 1596 bp mRNA linear HTC 03-APR-2004
 DEFINITION Mus musculus 13 days embryo liver cDNA, RIKEN full-length enriched library, clone:2500002N04 product: hypothetical Aminoacyl-transfer RNA synthetases class-II/Dihydrodipicolinate synthetase containing protein, full insert sequence.
 ACCESSION AK010857
 VERSION AK010857.1 GI:12846588
 KEYWORDS HTC; CAP trapper.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE ①
 AUTHORS Carninci, P. and Hayashizaki, Y.
 TITLE High-efficiency full-length cDNA cloning
 JOURNAL Meth. Enzymol. 303, 19-44 (1999)
 MEDLINE 99279253
 PUBMED 10349636
 REFERENCE ②
 AUTHORS Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.
 TITLE Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
 JOURNAL Genome Res. 10 (10), 1617-1630 (2000)
 MEDLINE 20499374
 PUBMED 11042159

FEATURES
 source Location/Qualifiers
 1. 1596
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="FANTOM DB:2500002N04"
 /db_xref="taxon:10090"
 /clone="2500002N04"
 /tissue_type="liver"
 /clone_lib="RIKEN full-length enriched mouse cDNA library"
 /dev_stage="13 days embryo"
 CDS
 28. 1020
 /note="unnamed protein product; hypothetical Aminoacyl-transfer RNA synthetases class-II/Dihydrodipicolinate synthetase containing protein (InterPro|IPR002106, InterPro|IPR002220, evidence: InterPro) putative"
 /codon_start=1
 /protein_id="BAB27226.1"
 /db_xref="GI:12846589"
 /translation="MLGPQIWASMRQGLSRGLSRNVKGMKVDIAGIYPPVTPPTATA EVDYGGLEENLNLRTAFPPRGFVVQGSTGEFPFLTSLERLLEVSRVQAIPKDKFLIA GSGCESTQATVEMTVSMAQVGADVAMVVTPCYRGRMSSAALIHHTKVADVSPVIV LYSVPANTGLELPVDAVVTLSQHPNIIIGLKDSGGDVTRIGLIVHKTQDFQVLAGSA GFLASYAVGAVGGICGLANVLGAQVCQLERLCTGQWEEAQLQRLIEPKHCGDPA LWNTRAEENHGLVLLWRLSLRPVAGAEPRGGGTALGFQQQLALMTSRRRLA"

ORIGIN
 Query Match 89.3%; Score 908.6; DB 3; Length 1596;
 Best Local Similarity 94.5%; Pred. No. 4.2e-231;
 Matches 952; Conservative 0; Mismatches 54; Indels 1; Gaps 1;

Qy	9	ATGCTGGGCCCCCAAATCTGGGCTCCATGAGGCAGGGGCTGAGCAGGGGCTTGCTTAGG	68
Db	28	ATGCTGGGCCCCCAAATTTGGGCTCCATGAGGCAGGGTCTGAGCAGGGGCTTGCTTAGG	87
Qy	69	AACGTGAAGGGGAAGAAGATAGACATTGCCGGCATCTACCCACCCGTGACCACCCCATTC	128
Db	88	AATGTGAAGGGCATGAAGGTAGACATTGCCGGCATCTACCCACCCGTGACCACCCCATTC	147
Qy	129	ACCGCCACCGCAGAGTAGACTATGGGAACTGGAAGAGAACTGAACAACTGGCCGCC	188
Db	148	ACCGCCACCGCAGAGGTAGACTATGGGAACTGGAAGAGAACTGAACAGACTGGCCACC	207
Qy	189	TTCCCCTTTTCGAGGCTTCGTGGTCCAGGGCTCTACTGGAGAGTTTCCATTCTGACCAGC	248
Db	208	TTCCCCTTTTCGAGGCTTCGTGGTCCAGGGTTCGACTGGAGAGTTTCCGTTCTGACCAGC	267
Qy	249	CTTGAGCGCCTAGAGGTGGTGGAGCCGAGTGCAGCCAGGCCATACCAAGGACAAGCTCCTG	308
Db	268	CTCGAGCGCCTGGAGGTGGTGGAGCCGAGTGCAGCCAGGCCATACCAAGGACAAGTTCCTG	327
Qy	309	ATAGCCGGCTCTGGCTGCGAGTCCACGCAAGCCACAGTAGAGATGACTGTGAGCATGGCT	368
Db	328	ATAGCCGGCTCTGGCTGCGAGTCCACGCAAGCCACAGTAGAGATGACTGTGAGCATGGCT	387
Qy	369	CAGGTGGGTGCTGATGCCGCATGGTGGTGACCCCTTGTTACTATCGCGGCCGCATGAAC	428
Db	388	CAGGTGGGTGCTGATGTCGCATGGTGGTGACCCCTTGTTACTATCGTGGCCGCATGAGC	447

Qy	429	AGCGCTGCCCTCATTACCACTACACCAAGGTTGCTGATCTTTCTCCAATCCCGGTGGTG	488
Db	448	AGCGCTGCCCTCATTACCACTACACCAAGGTTGCTGACGTTTCTCCAATCCCTGTGGTG	507
Qy	489	CTGTACAGTGTCCCAGGCAACACGGGTCTAGAGCTGCCTGTGGATGCCGTGGTCACATTG	548
Db	508	TTGTACAGTGTCCCAGCCAATACGGGGCTAGAGCTACCTGTGGATGCCGTGGTTACATTG	567
Qy	549	TCTCAGCACCCAAATATCATTGGCTTGAAGGACAGTGGTGGAGATGTGACCAGGACTGGG	608
Db	568	TCTCAGCACCCAAATATCATCGGCTTGAAGGACAGTGGTGGAGATGTGACCAGGATTGGA	627
Qy	609	CTGATTGTTACAAGACCAGCAAGCAGGATTTCCAGGTGTTGGCTGGGTGAGTTGGCTTC	668
Db	628	CTGATAGTTACAAGACCAGCAAGCAGGATTTCCAGGTGTTGGCTGGGTGAGTTGGCTTC	687
Qy	669	CTCCTGGCCAGCTATGCTGTGGGAGCTGTTGGGGGCATATGTGGCCTGGCCAATGTCTTG	728
Db	688	CTCCTGGCCAGCTATGCTGTGGGAGCTGTTGGGGGCATATGTGGCTGGCCAATGTCTTG	747
Qy	729	GGGGCCCAGGTGTGCCAGCTGGAGAGACTCTGCCTCACAGGGCAGGGGGAAGCTGCCCAG	788
Db	748	GGGGCCCAGGTGTGCCAGCTGGAGAGACTCTGCCTCACAGGGCAGTGGGAAGCTGCCCAG	807
Qy	789	AGACTGCAGCACCGCCTCATCGAGCCC-AACACTGCGGTGACCCGGCGCTTTGGAATACC	847
Db	808	GAAGTACAGCACCGTCTCATCGAGCCCAAACACTGCGGTGACCCGGCGCTTTGGAATACC	867
Qy	848	AGGGCTGAAGAAAACCATGGACTGGTTTGGCTACTATGGAGGTCCCTGCCGTGCCCCCTT	907
Db	868	AGGGCTGAAGAAAACCATGGACTGGTTTGGCTACTATGGAGGTCCCTGCCGTGCCCCGTT	927
Qy	908	GCAGGAGTTGAGCCCCCTCAGAGGAAGAGGCGCTTCGCTTGGATTTAGCAACAATGGCTG	967
Db	928	GCAGGAGCTGAGCCCCACAGAGGAGGAGGCACTGCGCTTGGATTTAGCAACAATGGCTG	987
Qy	968	GCTTTAATGACAAGCGGGGACACCTGGTCTGAGCTGTCTCAGAATT	1014
Db	988	GCTTTAATGACAAGCAGGAGACGCTGGCCTGAGCTATCTGGGACTT	1034